UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/752,432	01/07/2004	Jason G. Maxham	53217-014	3709

7590   12/21/2004

MCDERMOTT, WILL & EMERY
600 13th Street, N.W.
Washington, DC 20005-3096

EXAMINER
LIANG, GWEN

ART UNIT	PAPER NUMBER
2162	

DATE MAILED: 12/21/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

		Application No.	Applicant(s)
Office Action Summary		10/752,432	MAXHAM ET AL.
		Examiner	Art Unit
		GWEN LIANG	2162

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on *17 September 2004*.

2a) ☐ This action is **FINAL**.   2b) ☒ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) *1-34* is/are pending in the application.

    4a) Of the above claim(s) *14-28* is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) *1-13 and 29-34* is/are rejected.

7) ☐ Claim(s) _____ is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All   b) ☐ Some * c) ☐ None of:

      1. ☐ Certified copies of the priority documents have been received.

      2. ☐ Certified copies of the priority documents have been received in Application No. _____.

      3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Election/Restrictions

1.   Applicant's election without traverse of Species I including claims 1-13, 29-34 in the reply filed on 9/17/2004 is acknowledged.

2.   Claims 14-24 of Species II, and claims 25-28 of Species III are withdrawn from further consideration as being drawn to nonelected inventions.

Claim Objections

3.   Claim 34 is objected to because of the following informalities:

With regard to independent claim 34, the use of "files extension", in line 4, is an improper use of English language and The Examiner proposes to amend the claim language to read "file extension".

Claim Rejections - 35 USC § 112

4.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5.   Claim 34 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

With regard to claim 34, the claimed subject matter "that is converted to ASCII text", in lines 4-5, renders the claim indefinite because it is unclear whether the clause "that is converted to ASCII text " is describing "extension" or "document".

With regard to claim 34, the claimed subject matter "that are indexed for web-based retrieval", in line 7, renders the claim indefinite because it is unclear what is being indexed. Is it "documents", "meta-data, text", "attachments", or "retrieval"?

With regard to claim 34, the claimed subject matter "said identification", in line 7, renders the claim indefinite because it is unclear whether "said identification" refers to "identification tag" (line 4), "identification number (line 5), or "identified for retrieval" (line 6).

Claim 34 recites the limitation "the cluster database" in line 7 of the claim. There is insufficient antecedent basis for this limitation in the claim. The limitation " the cluster database" is not found in any of the preceding features in the claim.

Claim Rejections - 35 USC § 103

6.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7.   Claims 1, 4-13, and 29-33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), and further in view of Getchius et al., "Getchius " (U.S. Patent No. 6,493,721).

With respect to claim 1, Burrows discloses a method ...comprising:

a)   determining a file type for each native document of the plurality of native documents (col. 7, lines 58-65, "For example, the page 200 of FIG. 4 can have

associated page attributes 250. Page attributes 250 can include □ADDRESS□ 251,

□DESCRIPTION□ 252, □SIZE□ 253, □DATE□ 254, □FINGERPRINT□ 255, □TYPE□

256, and □END_PAGE□ 257, for example. The symbol "□," represents one or more

characters which cannot be confused with the characters normally found in words, for

example "space," "underscore," and "space" (sp_sp)"; col. 8, lines 24-25, "The TYPE

attribute 256 may distinguish pages having different multimedia content or formatting

characteristics"; Figure 4, element 256);

      b)   creating a fingerprint for each native document (col. 8, lines 16-23, "The

FINGERPRINT 255 represents the entire content of the page. The fingerprint 255 can

be produced by applying one-way polynomial functions to the digitized content.

Typically, the fingerprint is expressed as an integer value. Fingerprinting techniques

ensure that duplicate pages having identical content have identical fingerprints. With

very high probabilities, pages containing different content will have different

fingerprints.");

      c)   de-duplicating each native document in accordance with the fingerprint

(col. 1, lines 42-45, Therefore, it is desired to provide a technique which minimizes the

likelihood that duplicate pages are indexed. The technique should also allow for

reindexing as duplicate pages are deleted."; col. 2, lines 41-42, "FIG. 24 shows a

process for detecting duplicate pages; FIG. 25 is a flow diagram of a process for

deleting pages;"; col. 5, lines 12-14, "The maintenance module 80 also effectively deals

with duplicate Web pages containing substantially identical content.");

    d)   extracting data from each native document; e) associating extracted data with a corresponding native document (col. 5, lines 33-38, "A page 200 can be defined as a data record including a collection of portions of information or "words" having a common database address, e.g., a URL. This means that a page can effectively be a data record of any size, from a single word, to many words, e.g., a large document, a data file, a book, a program, or a sequence of images."; col. 11, line 66 – col. 12, line 7, "The samples are used to generate summary entries 925 in the second level summary data structure 72. Each summary entry 925 includes the word 926 associated with the sample, and the sampled location associated with the word. In addition, the summary entry 925 includes a pointer 928 of the next entry in the compressed data structure 71 following the sampled entry. The summary data structure 72 can also be mapped into fixed size blocks or disk files to fully populate the summary data structure 72.", wherein the words are extracted from the native documents); and

    f)   distributing the plurality of native documents and extracted data amongst a plurality of nodes of the document management computer system (col. 1, lines 65-67, "FIG. 1 is a block diagram of a distributed database storing multimedia information indexed and searched according to the invention;").

However Burrows does not explicitly disclose that the distribution of the plurality of native documents and extracted data amongst a plurality of nodes is substantially equal.

Getchius teaches a method comprising distributing data substantially equally amongst a plurality of nodes (Abstract, "The system for performing online data queries

is a distributed computer system with a plurality of server nodes each filly redundant

and capable of processing a user query request.").

It would have been obvious to one having ordinary skill in the art at the time the

invention was made to incorporate a method of distributing data substantially equally

amongst a plurality of nodes as disclosed by Getchius into the method of distributing the

plurality of native documents and extracted data amongst a plurality of nodes of the

document management computer system as disclosed in Burrows so that each node is

capable of responding to any search request (col. 18, lines 41-43). One of ordinary skill

in the art would be motivated to make the aforementioned combination with reasonable

expectation of success.

Claim 4 is rejected for the reasons set forth hereinabove for claim 1 and

furthermore Burrows discloses a method wherein step (c) further comprises comparing

the fingerprint of each native document with a plurality of fingerprints comprised of the

fingerprints for each native document to be uploaded (col. 28, lines 40-47).

Claim 5 is rejected for the reasons set forth hereinabove for claim 1 and

furthermore Burrows discloses a method wherein step (c) further comprises comparing

the fingerprint of each native document with at least one fingerprint corresponding to a

native document stored in the document management computer system (col. 28, lines

40-47).

Claim 6 is rejected for the reasons set forth hereinabove for claim 4 and

furthermore Burrows discloses a method comprising discarding native documents that

are determined to be the same in accordance with the comparison of fingerprints (Title;

col. 1, lines 42-45; col. 8, lines 16-23).

Claim 7 is rejected for the reasons set forth hereinabove for claim 5 and

furthermore Burrows discloses a method comprising discarding native documents that

are determined to be the same in accordance with the comparison of fingerprints (Title;

col. 1, lines 42-45; col. 8, lines 16-23).

Claim 8 is rejected for the reasons set forth hereinabove for claim 1 and

furthermore Burrows discloses a method wherein step (d) further comprises creating at

least one data file corresponding to the extracted data for each native document (col.

11, line 66 – col. 12, line 7).

Claim 9 is rejected for the reasons set forth hereinabove for claim 1 and

furthermore Burrows discloses a method wherein step (d) further comprises creating a

plurality of data files corresponding to the extracted data for each native document (col.

11, line 66 – col. 12, line 7).

Claim 10 is rejected for the reasons set forth hereinabove for claim 9 and

furthermore Burrows discloses a method wherein the plurality of data files includes files

selected from a group consisting of a text file, a meta data file, an XML file and a HTML

file (col. 8, line 66 – col. 9, line 8).

Claim 11 is rejected for the reasons set forth hereinabove for claim 10 and

furthermore Burrows discloses a method wherein in step (e), a data table is created for

at least one native document for defining an association with the plurality of data files

(col. 14, lines 35-40).

Claim 12 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein in step (e), a data table is created for at least one native document for defining an association with extracted data (col. 14, lines 35-40).

Claim 13 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Getchius discloses a program product, comprising executable code transportable by at least one machine readable medium, wherein execution of the code by at least one programmable computer causes the at least one programmable computer to perform a sequence of steps, comprising the steps recited in claim 1 (col. 19, line 52 – col. 20, line 8).

Claim 29 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a system comprising a computer in communication with the plurality of computer nodes for receiving a plurality of input files to be uploaded to the plurality of computer nodes (col. 2, lines 51-56, "FIG. 1 shows a distributed computer system 100 including a database to be indexed. The distributed system 100 includes client computers 110 connected to server computers (sites) 120 via a network 130. The network 130 can use Internet communications protocols (IP) to allow the clients 110 to communicate with the servers 120.").

The subject matter of claims 30 and 33 are rejected in the analysis above in claim 1, and therefore these claims are rejected on that basis..

The subject matter of claims 31 and 32 are rejected in the analysis above in claims 8 and 10 respectively, and therefore these claims are rejected on that basis.

8.   Claim 2 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows

(U.S. Patent No. 5,745,900), further in view of Getchius et al., "Getchius " (U.S. Patent

No. 6,493,721), and further in view of Okabe et al., "Okabe " (U.S. Publication No.

2001/0025287)

Claim 2 is rejected for the reasons set forth hereinabove for claim 1. However

the combination of Burrows and Getchius does not explicitly teach a method comprising

the step of extracting native document(s) included in the plurality of documents from an

archive file.

Okabe teaches the step of extracting native document(s) included in the plurality

of documents from an archive file (page 6, section [0077]).

It would have been obvious to one having ordinary skill in the art at the time the

invention was made to incorporate a step of extracting native document(s) included in

the plurality of documents from an archive file as disclosed by Okabe into the method of

managing a plurality of native documents as disclosed in the combination of Burrows

and Getchius. The motivation obviously is to obtain documents from the archive

through the extraction (page 6, section [0077]). One of ordinary skill in the art would be

motivated to make the aforementioned combination with reasonable expectation of

success.

9.   Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows

(U.S. Patent No. 5,745,900), further in view of Getchius et al., "Getchius " (U.S. Patent

No. 6,493,721), . and further in view of Zabetian (U.S. Publication No. 2001/0011350).

Claim 3 is rejected for the reasons set forth hereinabove for claim 1. However the combination of Burrows and Getchius does not explicitly teach a method wherein the fingerprint for each native document is created using a MD5 checksum.

Zabetian teaches a method wherein the fingerprint for each native document is created using a MD5 checksum (page 4, section [0037]).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a method wherein the fingerprint for each native document is created using a MD5 checksum as disclosed by Zabetian into the method of creating a fingerprint for each native document as disclosed in the combination of Burrows and Getchius, where a tamper proof checksum algorithm is desired, MD5 with DES encryption can be used (MD5-DES) (page 4, section [0037]). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

10.   Claim 34 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), and further in view of Dombrowski et al., "Dombrowski " (U.S. Patent No. 6,233,631).

With respect to claim 34, Burrows discloses a system ...comprising:

a PC type computer connected in a parallel cluster (col. 2, lines 51-56),

said computer using an operating system that stores electronic documents in a hard disk drive throughout the cluster (col. 3, lines 1-4 & 34-44; col. 11, lines 38-44; col. 15; lines 11-14, "This would be the case where the database indexed, the client

programs, the search engine 140, and the index 70 all reside on a single computer

system, e.g., a PC or workstation."),

said operating system defining a document identification tag (col. 8, lines 16-23,

"The FINGERPRINT 255 represents the entire content of the page. The fingerprint 255

can be produced by applying one-way polynomial functions to the digitized content.

Typically, the fingerprint is expressed as an integer value. Fingerprinting techniques

ensure that duplicate pages having identical content have identical fingerprints. With

very high probabilities, pages containing different content will have different

fingerprints.");

where each document is identified by its files extension (col. 7, lines 58-65, "For

example, the page 200 of FIG. 4 can have associated page attributes 250. Page

attributes 250 can include ☐ADDRESS☐ 251, ☐DESCRIPTION☐ 252, ☐SIZE☐ 253,

☐DATE☐ 254, ☐FINGERPRINT☐ 255, ☐TYPE☐ 256, and ☐END_PAGE☐ 257, for

example. The symbol "☐," represents one or more characters which cannot be

confused with the characters normally found in words, for example "space,"

"underscore," and "space" (sp_sp)"; col. 8, lines 24-25, "The TYPE attribute 256 may

distinguish pages having different multimedia content or formatting characteristics";

Figure 4, element 256);

and given a unique identification number (col. 26, lines 4-6, "Each entry 2201

includes an identification (page_id) 2210 of a qualified page"),

each of a plurality of documents having at least one of either meta-data, text or

attachments identified for retrieval that are indexed for web-based retrieval from the

cluster database (col. 8, line 66 – col. 9, line 8, "Attribute values or metawords can be

generated for portions of a page. For example, the words of the field 230 may be the

"title" of the page 200. In this case the "title" has a first word 231 and a last word 239.

In "html" pages, the titles can be expressly noted. In other types of text, the title may be

deduced from the relative placement of the words on the page, for example, first line

centered. For titles, the parsing module 30 can generate a ☐BEGIN_TITLE☐ pair and

an ☐END_TITLE☐ pair to be respectively associated with the locations of the first and

last words of the title."),

said identification of the plurality of documents forming a cluster data base that is

web-searchable by use of a predetermined descriptive term (col. 3, lines 28-33, "In

order to identify pages of interest among the millions of pages which are available on

the Web, a search engine 140 is provided. The search engine 140 includes means for

parsing the pages, means for indexing the parsed pages, means for searching the

index, and means for presenting information about the pages 200 located.").

However Burrows does not explicitly teach a system where each document is

converted to ASCII text.

Dombrowski teaches a system where each document is converted to ASCII text

(Abstract; col. 1, "The computer converts the machine usage data into a format

compatible for the given user, for example, converting to a generic ASCII type file

format or to an MS Excel format."; lines 47-51, "To overcome this deficiency in the prior

art, it would be desirable to offer the customer an opportunity to transmit data directly to

PC's and then convert the data into a generic ASCII type file format that may be easily integrated into a preferred accounting system").

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a system where each document is converted to ASCII text as disclosed by Dombrowski into the an electronic document management system as disclosed in Burrows, to provide machine usage data to a computer for a given user by setting the machine for transmission of the machine usage data to the computer and transmitting the machine usage data to the computer for various levels of usage. The computer converts the machine usage data into a format compatible for the given user, for example, converting to a generic ASCII type file format (Abstract). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to GWEN LIANG whose telephone number is 571-272-4038. The examiner can normally be reached on 9:00 A.M. - 5:30 P.M. Monday and Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JOHN BREENE can be reached on 571-272-4107. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

G.L.
13 December 2004

SHAHID ALAM
PRIMARY EXAMINER